PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04Q 7/38, H04B 7/185, H04J 3/12 — A1

(11) International Publication Number: WO 99/27745

(43) International Publication Date: 3 June 1999 (03.06.99)

(21) International Application Number: PCT/US98/25033

(22) International Filing Date: 23 November 1998 (23.11.98)

(30) Priority Data:
09/976,697 24 November 1997 (24.11.97) US

(71) Applicant: ERICSSON INC. [US/US]; 7001 Development Drive, P.O. Box 13969, Research Triangle Park, NC 27709 (US).

(72) Inventors: JOHNSON, Phillip, Marc; 2830 Crestscene Trail, Raleigh, NC 27603 (US). GREGORY, David; 1703 Laughridge Drive, Cary, NC 27511 (US). WAN, Yongbing; 33 Vetrina, Irvine, CA 92606 (US).

(74) Agents: MOORE, Stanley, R. et al.; Jenkens & Gilchrist, P.C., Suite 3200, 1445 Ross Avenue, Dallas, TX 75202 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

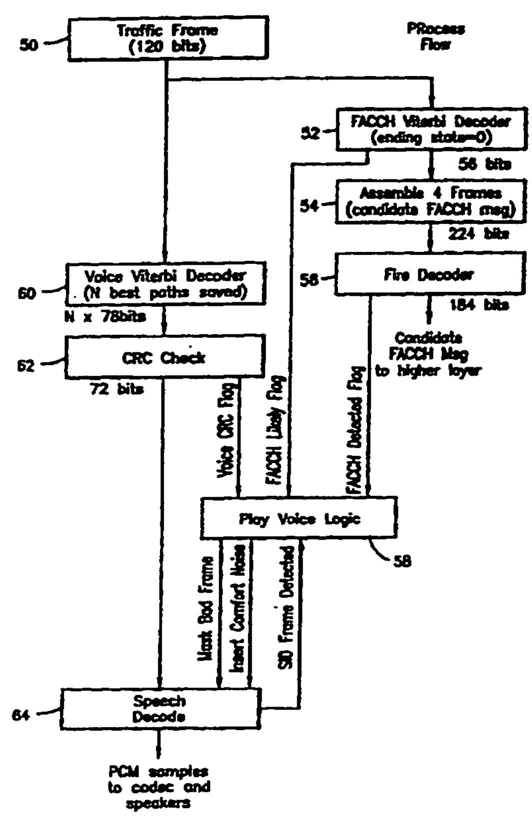
Published

With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SYSTEM AND METHOD FOR DETECTING SPEECH TRANSMISSIONS IN THE PRESENCE OF CONTROL SIGNALING

(57) Abstract

A telecommunications system and method for improving the detection of speech and control signals within a telecommunications transmission, particularly, reducing the probability that the control signals and other non-speech transmission segments are interpreted as speech and played. Also, the system and method of the present invention is directed to techniques for reducing the probability that random noise during discontinuous transmission periods are interpreted as speech and played.



SYSTEM AND METHOD FOR DETECTING SPEECH TRANSMISSIONS IN THE PRESENCE OF CONTROL SIGNALING

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to a communications system and method, particularly, to a communications protocol for the detection of speech transmissions amid control signals, and, more particularly, to an improved system and method for distinguishing valid speech frame transmissions from control signals and random radio/frequency (RF) noise, thereby avoiding speech quality degradation by minimizing the chance of incorrectly processing a non-speech frame as if it were speech.

Background and Objects of the Invention

The evolution of wireless communication over the past century, since Guglielmo Marconi's 1897 demonstration of radio's ability to provide continuous contact with ships sailing the English Channel, has been remarkable. Since Marconi's discovery, new wireline and wireless communication methods, services and standards have been adopted by people throughout the world. This evolution has been accelerating, particularly over the last ten years, during which time the mobile radio communications industry has grown by orders of magnitude, fueled by numerous technological advances that have made portable radio equipment smaller, cheaper and more reliable. The exponential growth of mobile telephony will continue in the coming decades, as this wireless network interacts with and eventually overtakes the existing wireline networks.

The Global System for Mobile (GSM) communications is a second generation cellular system standard developed to solve various fragmentation problems of the first cellular systems in Europe. GSM is the world's first cellular system to specify digital modulation and network level architectures and services. Currently, GSM is the most popular standard for new radio and personal communications equipment throughout the world.

time prompting the transmitter to resume normal transmission. Accordingly, the receiving radio must always be ready to receive speech. This implies that the receiver remains on during DTX periods searching for a valid speech frame. There is a chance that the random noise on the air will occasionally pass through the receiver and be interpreted as a valid speech frame which gets played. Without some corrective action (as described in this disclosure), the mathematical probability of a noise frame passing into the audio path during a DTX period is quite significant. If a frame of random noise does mistakenly get passed to the speech decoder and played, it will likely create a pop or other audio artifact within the DTX period, thereby degrading the perceived audio quality.

In an effort to prevent the aforementioned sources of audio degradation, current digital standards have some reasonably straightforward and robust methods for distinguishing speech and FACCH signals. Also, DTX periods are currently distinguished by using the quality of a Viterbi metric or the strength of sync correlation, as is understood in the art. The problem is that the SAIS is presently inadequate to prevent these sources of audio degradation.

Accordingly, it is an object of the present invention to prevent the interpretation of FACCH or other overriding control messages as speech, thereby avoiding artifacts that degrade speech quality.

It is another object of the present invention to avoid the conversion of random noise into speech frames during DTX periods.

SUMMARY OF THE INVENTION

The present invention is directed to a communications system and method for improving the detection of speech frames within a telecommunications transmission, particularly, reducing the probability that control signals get interpreted as speech frames and played as audio. Also, the system and method of the present invention is directed to techniques for reducing the probability that random RF noise gets interpreted as speech frames and played as audio.

an ACeS system, it is useful to first describe the communication environment of the GSM system upon which ACeS is based, as well as other environments where control signals are interspersed with speech data.

Under GSM, speech data and control signal data from the Fast Associated Control Channel (FACCH) are transmitted over a multiplicity of bursts. The format of a normal transmission burst is shown in FIGURE 1. With speech processing at the rate of 13 Kbps, 260 bits of speech are generated every 20 ms. With block and convolutional coding, those 260 bits are expanded to 456 bits for each 20 ms frames of speech. The 456 bits are divided into four 114 bit blocks, each of which are mapped to the data fields D_1 and D_2 shown in FIGURE 1.

The 42.25 additional bits in the burst include: a 26-bit training sequence for the equalizer, i.e., (SYNC) bits, allowing burst demodulation with no information from previous bursts; time slot start (S) and end (E) tail flags of 3 bits each, allowing the impulse response of the channel and modulation filter to terminate within the burst, ensuring that end bit demodulation is the same as at the burst middle; two one-bit flags (F_1 and F_2) to distinguish speech from FACCH; and 8.25 guard bits (GB) for up/down ramping time. The F_1 bit indicates whether the data in the preceding burst was either speech data or FACCH data, and the F_2 bit indicates the origin of the data in the current burst.

With Time Division Multiple Access (TDMA), the aforementioned four blocks of 114 bits are assigned to a particular time slot (TS) within a frame FR, e.g., TS_2 in FIGURE 1. In GSM, each frame FR has eight timeslots (TS_0 to TS_7) therein, each of which is assigned to a different user. In turn, frame FR is one of 26 frames in a multiframe MF, as is understood in the art.

As discussed, FACCH messaging is implemented by replacing one 20 ms frame of speech data with one FACCH message. Although the number of significant FACCH bits are fewer, i.e., 184 bits, than that of speech data bits, FACCH control signals are encoded more heavily to preserve the integrity of the control message during transmission. After such encoding, the FACCH message is, like speech, 456 bits long. Instead of a traffic channel, however, the FACCH message is sent through a control channel, particularly, as part of the Associated Control Channel. Since both the traffic and control channels are logical channels sharing a common physical

Additionally, the bit error rate (BER) estimate from a FACCH Viterbi decode and from the speech Viterbi decode may be used, as is understood in the art.

Also, the Personal Digital Cellular (PDC) standard air interface defines a single "steal flag" in its slot structure. As with GSM, this mechanism is fairly robust. Additionally, as with D-AMPS, a CRC is defined for both speech and FACCH. Audio is, therefore, only played if the steal flag indicates that the current frame is speech, the FACCH CRC failed and the speech CRC passed.

Although similar to GSM in many ways, the ACeS system is designed to operate with much greater capacity. Because of the severe power and possible bandwidth limitations in a satellite communications system, speech must be coded at bit-rates much lower than those in GSM. Accordingly, instead of encoding speech at 13 Kbps, ACeS codes speech at 3.6 Kbps, which is equivalent to 72 bits per 20 ms, which becomes 120 bits in basic mode after channel encoding.

A representative diagram of a satellite-cellular communication network is shown in FIGURE 3. A satellite 10, such as one in geostationary orbit over SouthEast Asia in the ACeS system, forwards and receives digital information to and from a variety of land-based equipment, such as a Network Control Center (NCC) 12 for controlling call management functions, a Land-Earth Station (LES) 14 and a plurality of cellular phones 16. The LES 14, a mobile switching center/visitor location register (MSC/VLR) 18 and an interworking unit 20 handle the traffic channels, as is understood in the art.

Through the interworking unit 20, cellular communications are also accessible through a public Switched Telephone Network (PSTN) 22 to a facsimile 24, a regular non-cellular telephone 26 and a service computer 28 via a modem 30. Other cellular devices, such as other cellular phones 32, may also access the satellite through a cellular link 34.

The format of an ACeS burst is different from that of a GSM burst, as shown in FIGURE 4, and incorporates more data bits therein, i.e., 120 per burst (D_1 and D_2) as compared to 114 for GSM. The SYNC field has been shortened and the steal flag bits F_1 and F_2 have been eliminated in order to provide more data bits in the D_1 and D_2 fields. The SAIS suggests that speech should be processed whenever the speech CRC passes. As discussed, however, some FACCH and other anomalous signals may improperly pass the speech CRC, thereby degrading the speech quality.

suggested by the SAIS standard, will result in poor speech quality and loss of customer satisfaction.

Furthermore, the SAIS defines a DTX mode which is very similar to GSM's DTX mode. The speech coder includes a Voice Activity Detector (VAD). Whenever the VAD determines that voice is no longer active, a transmitter may enter DTX mode. When the transmitter enters such a mode, it ceases to transmit in every one of its assigned timeslots. Instead, it transmits at a lower rate (typically about once per second). The frames which are transmitted at this lower rate are different from normal speech frames. These special frames are termed "silence descriptor" (SID) frames. They characterize the acoustic background noise at the transmitter. The receiver may then use the SID frames to emulate any background noise at the transmitter. In the time between SID frame transmissions during a DTX period, the receiver is receiving nothing. Once voice activity resumes at the transmitter, the transmitter will exit the DTX period and begin transmitting normal voice frames again. Thus, the receiver must always be ready for the transmitter to exit the DTX period.

At the receiver, the periodic SID frames are used by the speech decoder to insert "comfort noise." During periods when valid SID frames are not being received, the noise characteristics of the last received SID frame are played. The speech decoder, however, must be ready to begin playing voice again when voice transmission restarts. During DTX periods, the transmitter is generally not transmitting any traffic frames to the receiver for long periods of time. However, the receiver is still demodulating whatever is on the air in anticipation of the resumption of speech. The random or "bad frame" data provided by the demodulator will occasionally (on the order of 1-10% of the time) create a CRC pass. Considering the length of typical DTX periods (on the order of hundreds of frames), it becomes very likely that random data during DTX periods will create a speech CRC pass. As noted, if any of this random data gets played as audio, it is likely to create degrading artifacts within the comfort noise. This bad frame will probably be followed by random data during the DTX period which may be interpreted (correctly) as bad frames. This will force frame repeats, effectively lengthening the period of time the misinterpreted bad frame will be played, causing further user annoyance.

In view of some convolutional coding peculiarities within the SAIS, convolutional coding and an implementation thereof will now be discussed. A

16 or LES 14) after the incoming signal exits the equalizer/demodulator therein. The frame of data (120 bits) after demodulation is represented in box 50. This data is fed both to a FACCH Viterbi decoder 52 and a voice Viterbi decoder 60. Within decoder 52, the 120 bits are Viterbi decoded to an output 56 bits and the trellis is forced to terminate in the zero state. (The traceback is always from the zero state.) If the zero state happened to have the best metric of all the ending states, a FACCH likely flag, discussed further herein, is set. The 56 bit frame is then passed to an assembler 54 which assembles the received frame of data with the three prior frames, the four of which are then sent to a fire decoder 56, which accepts a 224 (56 x 4) bit segment of data and outputs 184 bits after fire decoding. If the fire decoder 56 determines that a valid four frame FACCH message was received, a FACCH detected flag is set and passed to a play voice logic device 58, as also discussed further herein. The properly received and decoded FACCH message is then passed along to the appropriate higher layer for processing.

As within the FACCH Viterbi decoder 52, the voice Viterbi decoder 60 accepts the 120-bit traffic frame but outputs N candidate 78-bit frames. These N candidate frames are found by choosing the N ending states in the Viterbi trellis which have the best metrics. The N candidate frames are then forwarded to a CRC check 62 which attempts to find the best frame among the N candidate frames which has a passing CRC. If successful in finding such a frame, the check 62 sets a voice CRC flag, which is forwarded to the play voice logic device 58, and forwards 72 bits of speech data (6 bits were used in the CRC checking) to a speech decoder 64.

With reference now to FIGURE 7, there is illustrated some of the methodology of the play voice logic device 58, which implements many of the features of the present invention. As noted in FIGURE 6, the results of the three flags, i.e., the voice CRC flag from the CRC check 62, the FACCH detected flag from the fire decoder 56 (actually the inverted value thereof) and the FACCH likely flag from the FACCH Viterbi decoder 52 (actually the inverted value of the logical addition of the current and previous frames) are fed into an AND logical summation function (box 70).

If the summation result (box 72) of the aforementioned inputs is one (TRUE), then control is passed to box 74, indicating that the particular incoming frame of data is most likely speech; otherwise control is passed to box 86. At box 74, a good frame counter (GFC), where "good" means speech, is incremented and control is passed to

previous speech frame. However, it is assumed here that the bad frames are due to a brief impairment on the channel rather than the transmitter having entered a DTX period.

Through use of the play voice logic device 58, shown in FIGURE 6, with the aforedescribed logic flow therein, as shown in FIGURE 7, most of the previously discussed anomalous situations causing speech quality degradation are handled. For example, the logic shown in FIGURE 6 makes it unlikely that FACCH bursts will be mistakenly interpreted as speech and played out of the speech decoder, resulting in the aforementioned audio pops. With reference to FIGURE 5, if an incoming burst represents the fourth (and last) burst of a FACCH message, the fire decoder 56 should set the FACCH detected flag, forcing the speech decoder 64 to take corrective action, e.g., the speech decoder 64 upon receipt of a bad frame mask flag controls whether to frame repeat or insert comfort noise. Typically, the speech decoder 64 repeats up to four frames in a row, i.e., M=4, and then starts comfort noise insertion.

Regarding the more problematic previous three FACCH bursts, the methodology of the present invention assists in this determination also. If the FACCH Viterbi decoder 52 determines that two consecutive bursts have zero ending states which represent/contain the best metrics of all the ending states, it is likely that the particular incoming frame or burst is part of a FACCH message transmission. Here, the speech decoder must also take corrective action, as described, to mask these "bad" frames. Lastly, with only the more problematic first FACCH message burst which is still in doubt, and as a final precaution, the voice CRC check 62 for the incoming frame must pass before that frame is passed through the speech decoder 64. As before, if the voice CRC fails, the speech decoder 64 will be forced to take the aforedescribed corrective actions.

Also, through use of the play voice logic device 58 and associated circuitry therein, shown in FIGURES 6 and 7, anomalous situations arising out of DTX mode usage are addressed as well. For example, at the onset of a DTX period, the speech encoder at the transmitting end begins creating Silence Descriptor (SID) frames which may be used by the speech decoder 64 to determine the correct noise characteristics for CNI. The transmitter sends a limited number of these SID frames before the onset of the DTX period. Whenever the speech decoder 64 receives a SID frame, it begins CNI and sets a SID frame detected flag, which is available after the speech decoder 64

inputs and has a threshold to determine whether to take a good frame or bad frame path.

In another alternative embodiment of the present invention, a four frame block, e.g., EF_2 to EF_5 in FIGURE 5, could be fire decoded to determine if it was a FACCH message. If not, the oldest frame would then be speech decoded if the CRC passed. This embodiment, however, is not preferred because of the additional 60 ms of delay introduced.

It should be understood that although the aforedescribed preferred embodiment employs TDMA technology, the principles of the present invention are applicable to other access techniques, e.g., Code Division Multiple Access (CDMA) technology, TDMA/CDMA hybrids and any other digital telecommunications system employing speech frames.

While the invention has been described in connection with preferred embodiments thereof, it is to be understood that the scope of the invention is not limited to the described embodiments, but is intended to encompass various modifications and equivalents within the spirit and scope of the appended claims.

second voice decoder setting said speech flag if said decoded particular transmission frame decodes pursuant to a second metric.

6. The receiver apparatus according to claim 5, where said second voice transmission decoder is a cyclic redundancy code check, said speech flag set if said decoded particular transmission frame passes said cyclic redundancy code check.

7. The receiver apparatus according to claim 1, wherein said detector further comprises:

a control signal transmission decoder, said control signal transmission decoder receiving said particular transmission frame and setting said control signal likely flag if said particular transmission frame decodes pursuant to a third metric forming a candidate control signal frame.

8. The receiver apparatus according to claim 7, wherein said control signal transmission decoder is a Viterbi decoder, said control signal likely flag being set if said particular transmission frame decodes pursuant to said Viterbi decoder.

9. The receiver apparatus according to claim 7, further comprising:

a fire decoder, said fire decoder receiving said candidate control signal frame and a plurality of prior transmission frames, from said control signal transmission decoder, and setting said control signal detected flag if said fire decoder determines that a valid control signal transmission was received.

10. The receiver apparatus according to claim 9, further comprising an assembler, said assembler receiving said candidate control signal frame, assembling said candidate control signal frame with said plurality of prior transmission frames, forming an assembled frame group, and forwarding said assembled frame group to said fire decoder.

11. The receiver apparatus according to claim 10, wherein said assembler assembles four said frames, one being said candidate control signal frame and the remaining three being said prior transmission frames.

18. The receiver apparatus according to claim 1, wherein said control signal transmissions are Fast Associated Control Channel (FACCH) signals within said series of transmission frames.

19. The receiver apparatus according to claim 18, wherein said FACCH signals comprise four consecutive transmission frames in said series.

20. The receiver apparatus according to claim 1, wherein said telecommunications system is based upon Satellite Air Interface Specification protocols.

21. The receiver apparatus according to claim 1, wherein said receiver is within a mobile terminal in wireless communication with a base station.

22. The receiver apparatus according to claim 1, wherein said receiver is within a base station.

23. In a digital telecommunications system having a first communication system and a second communication system, the first and second communication systems coupled together by way of a communication channel, a combination with the first and second communication systems of communication circuitry for transmitting and receiving, respectively, a plurality of speech frames therebetween, said circuitry comprising:

transmission means within said first communication system, said transmission means generating and transmitting a substantially continuous series of transmission frames containing said speech frame segments therein across said communication channel, said transmission means also generating and transmitting a plurality of transmission frames of a control signal across said communication channel, said control signal having precedence over said speech and a plurality of control signal frames overriding a corresponding plurality of said speech frames;

reception means within said second communication system, said reception means for receiving said substantially continuous sequence of transmission

29. A digital telecommunications system having a transmitter and a receiver coupled together by way of a communication channel, a substantially continuous series of transmission frames containing speech and a plurality of control signals therein passing across said channel from said transmitter to said receiver across such channel, said control signal having precedence over and overriding said speech, said telecommunications system comprising:

a detector, attached to said receiver, for detecting said series of transmission frames, said detector setting a multiplicity of flags, said flags comprising a speech flag set if a particular transmission frame contains speech therein, a control signal detected flag if said particular transmission frame contains said control signals therein and a control signal likely flag if said particular transmission frame potentially contains said control signals therein; and

a summation device, attached to said detector, said detector applying said multiplicity of flags to said summation device, whereby speech transmissions play at said receiver whenever said summation device indicates a speech transmission.

30. The telecommunications system according to claim 29, wherein said receiver, detector and summation device are within a mobile terminal in wireless communication with said transmitter.

31. The telecommunications system according to claim 29, wherein said receiver, detector and summation device are within a base station.

32. The telecommunications system according to claim 29, wherein said control signal is a Fast Associated Control Channel signal.

33. The telecommunications system according to claim 29, wherein said telecommunications system is based upon Satellite Air Interface Specification protocols.

34. The telecommunications system according to claim 29, wherein said control signal likely flag is set if said detection means determines that a best candidate

39. The method according to claim 35, wherein said telecommunications system is based upon Satellite Air Interface Specification protocols.

40. The method according to claim 35, further comprising steps of:

calculating a best candidate frame metric for said particular transmission frame pursuant to a first metric;

determining if said best candidate frame metric is a zero state; and

setting said control likely flag if said best candidate from metric is said zero state.

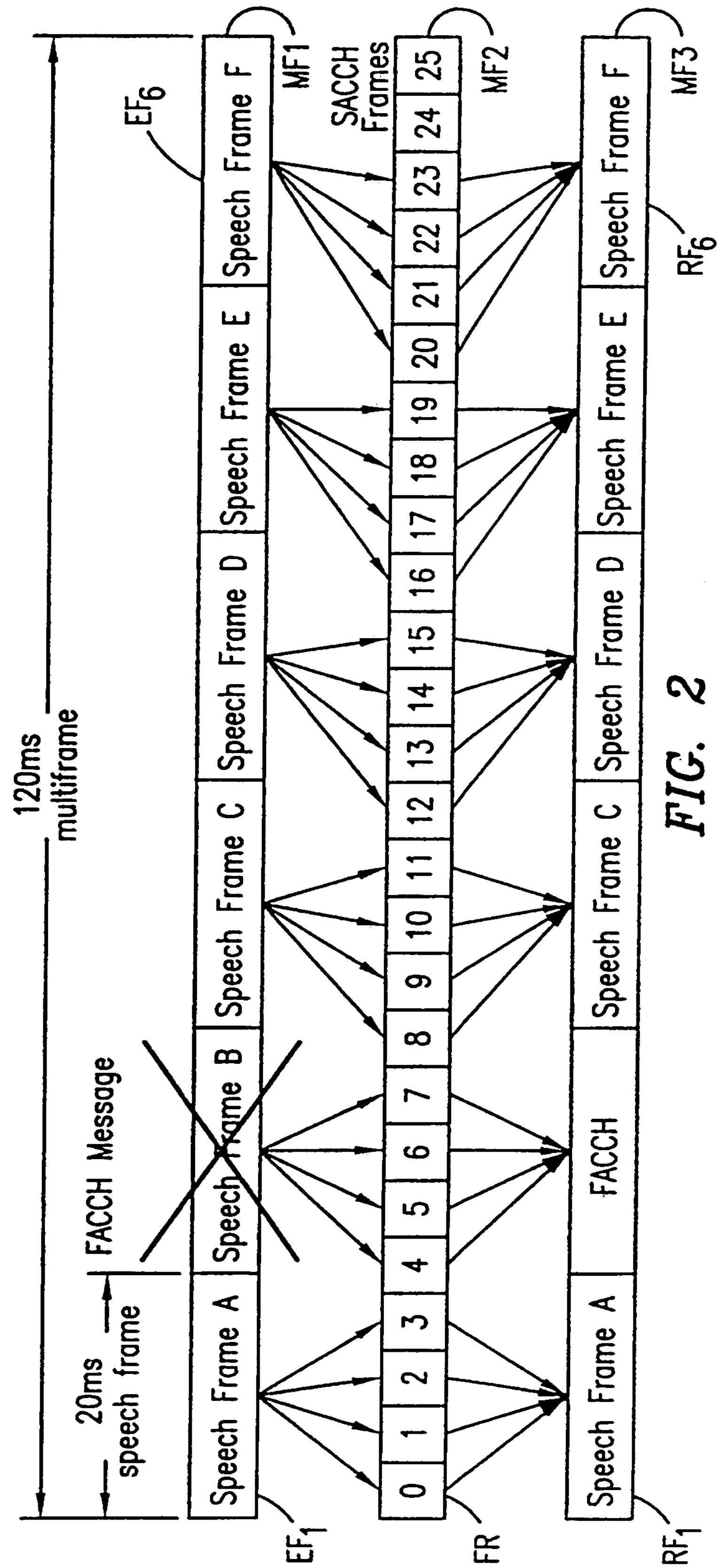

120ms multiframe
20ms speech frame
FACCH Message
Speech Frame A
Speech Frame B
Speech Frame C
Speech Frame D
Speech Frame E
Speech Frame F
EF1
EF6
MF1
SACCH Frames
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
FR
MF2
Speech Frame A
FACCH
Speech Frame C
Speech Frame D
Speech Frame E
Speech Frame F
RF1
RF6
MF3

FIG. 2

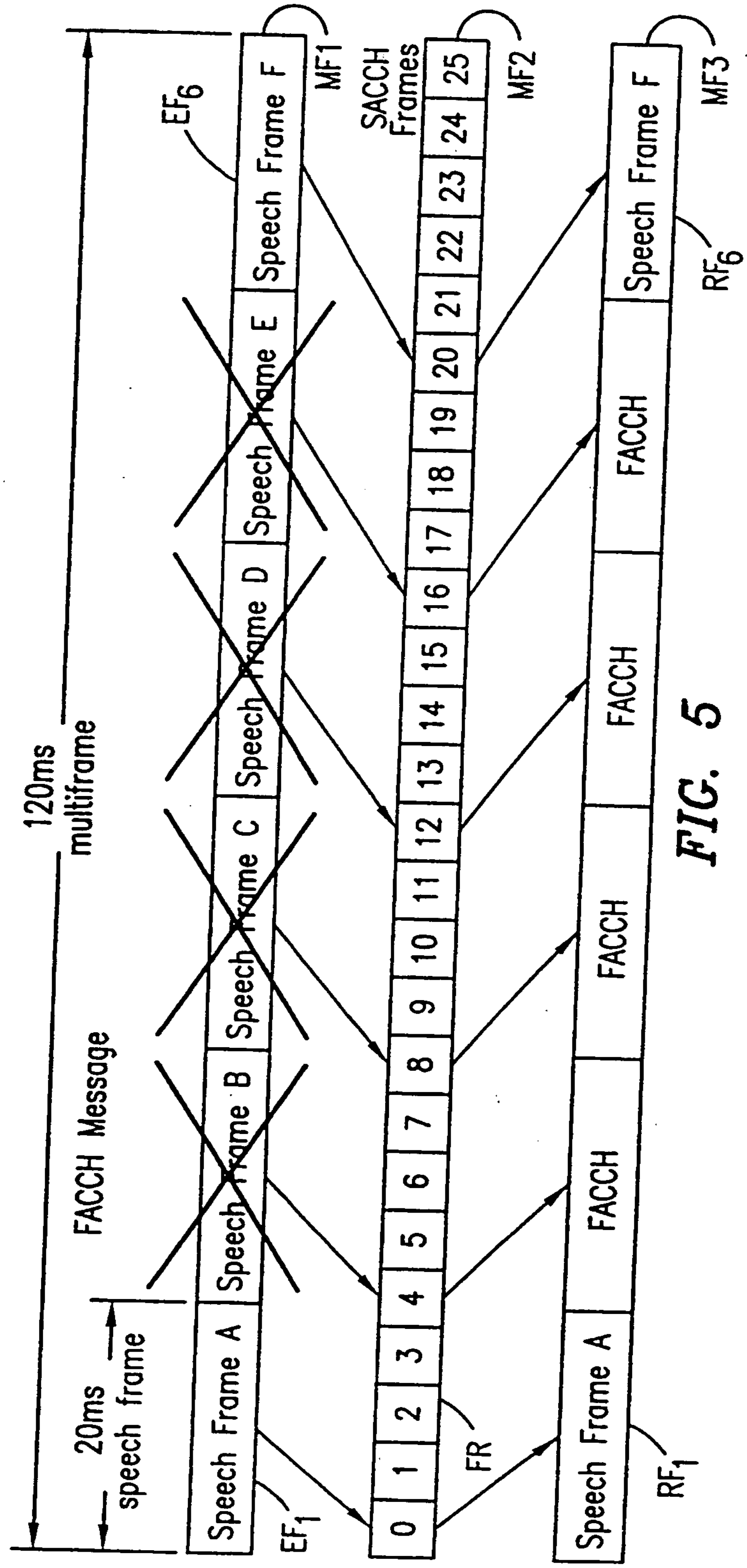

120ms multiframe
FACCH Message
20ms speech frame
Speech Frame A
Speech Frame B
Speech Frame C
Speech Frame D
Speech Frame E
Speech Frame F
EF_1
EF_6
MF1
SACCH Frames
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
MF2
FR
Speech Frame A
FACCH
FACCH
FACCH
FACCH
Speech Frame F
RF_1
RF_6
MF3

FIG. 5

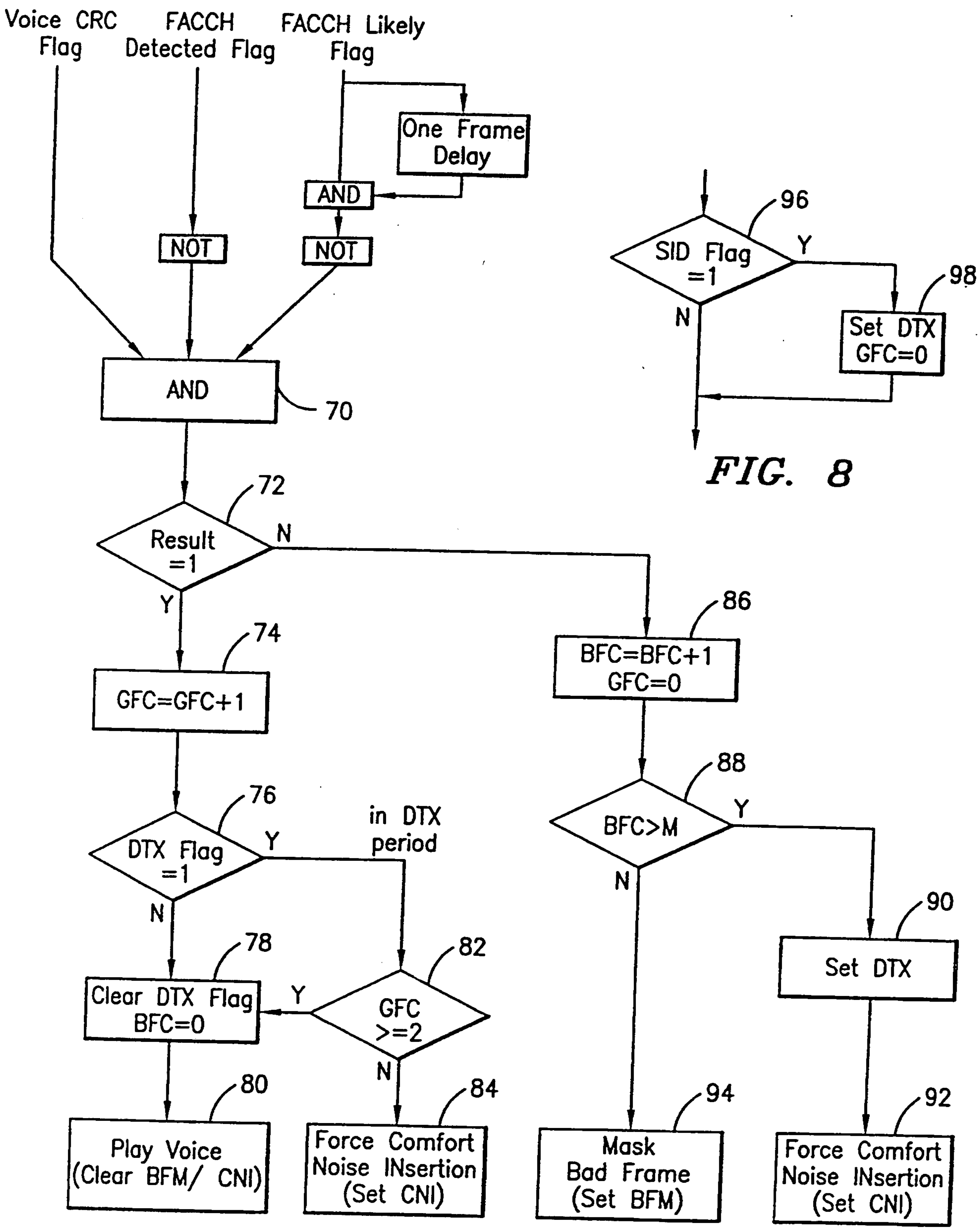

Voice CRC Flag
FACCH Detected Flag
FACCH Likely Flag
One Frame Delay
AND
NOT
NOT
AND
70
72
Result =1
N
Y
74
GFC=GFC+1
76
DTX Flag =1
Y
in DTX period
N
78
Clear DTX Flag BFC=0
Y
82
GFC >=2
N
80
Play Voice (Clear BFM/ CNI)
84
Force Comfort Noise INsertion (Set CNI)
86
BFC=BFC+1 GFC=0
88
BFC>M
Y
N
90
Set DTX
94
Mask Bad Frame (Set BFM)
92
Force Comfort Noise INsertion (Set CNI)

FIG. 7

96
SID Flag =1
Y
98
Set DTX GFC=0
N

FIG. 8

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 98/25033

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0651523 A	03-05-1995	JP 2601160 B	16-04-1997
		JP 7123462 A	12-05-1995
		US 5757851 A	26-05-1998
US 5499246 A	12-03-1996	GB 2260245 A	07-04-1993
		AU 647670 B	24-03-1994
		AU 2458592 A	08-04-1993
		DE 4233089 A	08-04-1993
		FR 2683412 A	07-05-1993
		IT 1255845 B	17-11-1995
		JP 5227231 A	03-09-1993
		SE 9202873 A	05-04-1993